ABSTRACT

Similarity Searching For Documents

5

     A method of searching a database, such as a database representative of content on the World Wide Web, to find documents similar to a query document, involves a step of decomposing 22 the query document 1 into elements of different data types. After this, for one or more of the elements in a first data

10
type, a first data type similarity search is conducted 23 to return match results from the database for the one or more elements in the first data type. For one or more of the elements in a second data type, a second data type similarity search is conducted to return match results from the database for the one or more elements in the first data type. The match results from the different data

15
types are combined with an appropriate weighting to provide query document match results. Data types can include text, picture and graphics, and also the layout of the overall document.